

Amendments to the Claims:

The following listing of claims will replace all prior versions, and listings, of claims in the application:

1. (Currently Amended) A document extracting ~~device~~ apparatus, comprising:
a document acquiring device to acquire a plurality of documents from an information source, according to a user-specific criteria, to be candidates for extraction;
a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing compute all degrees of similarity between the candidate plurality of documents, and express the degrees of similarity in a symmetric matrix;
and
a combination computing device to compute all combinations of any number of documents from the plurality of documents;
a sum of degrees of similarity computing device to compute, with respect to all of the combinations, a sum of the degrees of similarity between all of the documents that constitute each combination, based on all of the degrees of similarity expressed in the symmetric matrix; and
a document extracting device to extract a combination of documents constituting the combination with the smallest sum of the degrees of similarity among the plurality of documents constituting the respective combinations. from the candidate documents with a sum of the degrees of similarity between the candidate documents that is the smallest when any number of the combination of documents are extracted from among a group of the candidate documents.
2. (Currently Amended) The document extracting ~~device~~ apparatus according to Claim 1,

the similarity computing device comprising:

a character-string-dividing functional unit to divide each of the ~~eandidate~~
plurality of documents into predetermined character strings;

a character-string frequency computing functional unit to compute document
vectors of the ~~eandidate~~plurality of documents on the basis of a frequency of appearance of
the predetermined character strings divided by the character-string-dividing functional unit;
and

_____ a mutual similarity computing functional unit to compute the degrees of
similarity between the ~~eandidate~~plurality of documents on the basis of the document vectors
obtained from the character-string frequency computing functional unit.

3. (Currently Amended) The document extracting ~~device~~apparatus according to
Claim 2,

the character-string-dividing functional unit dividing each of the ~~eandidate~~
plurality of documents into predetermined character strings using any of the following
character string division methods: a morphological analysis method, an n-gram method, and a
stop-word method.

4. (Currently Amended) The document extracting ~~device~~apparatus according to
Claim 2,

the character-string frequency computing functional unit generating document
vectors obtained by weighting each of the ~~eandidate~~plurality of documents by a term
frequency and inverse document frequency (TFIDF) weighting method on the basis of a
frequency of appearance of the divided character strings.

5. (Currently Amended) The document extracting ~~device~~apparatus according to
Claim 2,

the mutual similarity computing functional unit computing degrees of similarity between the ~~candidate-plurality of~~ documents by a vector space method on the basis of the document vectors of the ~~candidate-plurality of~~ documents.

6. (Currently Amended) A computer-readable media having a document extracting program allowing a computer to serve as:

a document acquiring device to acquire a plurality of documents from an information source, according to a user-specific criteria, to be candidates for extraction;

a similarity computing device to compute all degrees of similarity between the plurality of documents, and express the degrees of similarity in a symmetric matrix;

a combination computing device to compute all combinations of any number of documents from the plurality of documents;

a sum of degrees of similarity computing device to compute, with respect to all of the combinations, a sum of the degrees of similarity between all of the documents that constitute each combination, based on all of the degrees of similarity expressed in the symmetric matrix; and

a document extracting device to extract documents constituting the combination with the smallest sum of the degrees of similarity among the plurality of documents constituting the respective combinations.

~~a similarity computing device to acquire a plurality of documents to be candidates for extraction and computing all degrees of similarity between the candidate documents; and~~

~~a document extracting device to extract a combination of documents from the candidate documents with a sum of the degrees of similarity between the candidate documents that is the smallest when any number of the combination of documents are extracted from among a group of the candidate documents.~~

7. (Currently Amended) The media according to Claim 6,
the similarity computing device comprising:
a character-string-dividing function to divide each of the ~~candidate~~ plurality of
documents into predetermined character strings;
a character-string frequency computing function to compute document vectors
of the ~~candidate~~ plurality of documents on the basis of a frequency of appearance of the
predetermined character strings divided by the character-string-dividing function; and
a mutual similarity computing function to compute the degrees of similarity
between the ~~candidate~~ plurality of documents on the basis of the document vectors obtained
by the character-string frequency computing function.

8. (Currently Amended) The media according to Claim 6,
the similarity computing device comprising:
a character-string-dividing function to divide each of the ~~candidate~~ plurality of
documents into character strings using any one of character string division methods;
a character-string frequency computing function to generate document vectors
obtained by weighting each of the documents by a term frequency and inverse document
frequency (TFIDF) weighting method on the basis of a frequency of appearance of the divided
character strings; and
a mutual similarity computing function to compute the degrees of similarity
between the ~~candidate~~ plurality of documents by a vector space method on the basis of the
document vectors of the ~~candidate~~ plurality of documents.

9. (Currently Amended) A document extracting method, comprising:
acquiring a plurality of documents from an information source, according to a
user-specific criteria, to be candidates for extraction;

computing all degrees of similarity between the plurality of documents, and
expressing the degrees of similarity in a symmetric matrix;

computing all combinations of any number of documents from the plurality of
documents;

computing, with respect to all of the combinations, a sum of the degrees of
similarity between all of the documents that constitute each combination, based on all of the
degrees of similarity expressed in the symmetric matrix; and

extracting documents constituting the combination with the smallest sum of
the degrees of similarity among the plurality of documents constituting the respective
combinations.

~~acquiring a plurality of documents to be candidates for extraction;~~
~~computing all degrees of similarity between the candidate documents; and~~
~~extracting a combination of documents from the candidate documents with a sum of the~~
~~degrees of similarity between the candidate documents that is the smallest when any number~~
~~of the combination of documents are extracted from among a group of the candidate~~
~~documents.~~

10. (Currently Amended) The document extracting method according to Claim 9,
further comprising:

dividing each of the documents into predetermined character strings,
computing a frequency of appearance of the divided character strings, computing document
vectors of the ~~candidate plurality of~~ documents on the basis of the frequency of appearance of
the predetermined character strings, and then computing the degrees of similarity between the
~~candidate plurality of~~ documents using the document vectors.

11. (Currently Amended) The document extracting method according to Claim 9,
further comprising:

dividing each of the ~~eandidate~~plurality of documents into predetermined character strings using any one of character string division methods, including a morphological analysis method, an n-gram method, and a stop-word method, computing document vectors of the ~~eandidate~~plurality of documents by weighting each of the documents by a term frequency and inverse document frequency (TFIDF) weighting method on the basis of a frequency of appearance of the divided predetermined character strings, and computing the degrees of similarity between the ~~eandidate~~plurality of documents using a vector space method on the basis of the document vectors.

12-14. (Canceled)